Science Journal of Public Health
2022; 10(3): 115-123
http://www.sciencepublishinggroup.com/j/sjph
doi: 10.11648/j.sjph.20221003.13
ISSN: 2328-7942 (Print); ISSN: 2328-7950 (Online)

Effects of Different Absenteeism Statistical Methods on Influenza Surveillance

Zhen Yang[1, 2], Cheng-hua Jiang[1, *]

[1]School of Medicine, Tongji University, Shanghai, China

[2]School of Medicine, Jinggangshan University, Jian, China

Email address:
34419502@qq.com (Zhen Yang), jchtongji@163.com (Cheng-hua Jiang)
[*]Corresponding author

To cite this article:
Zhen Yang, Cheng-hua Jiang. Effects of Different Absenteeism Statistical Methods on Influenza Surveillance. *Science Journal of Public Health.* Vol. 10, No. 3, 2022, pp. 115-123. doi: 10.11648/j.sjph.20221003.13

Received: May 1, 2022; **Accepted**: May 19, 2022; **Published**: May 26, 2022

Abstract: *Background* The existing absenteeism surveillance systems in China rely heavily on school doctors to collect data manually, but the low prevalence rate of school doctors makes it difficult to popularize this mode. The method of absenteeism statistics requires new breakthroughs. *Objective* The purpose of this study was to evaluate the feasibility of an established absenteeism surveillance system based on face recognition, and to explore the appropriate surveillance index for this system. *Methods* A primary school of about 1900 students was selected. Absenteeisms reported by school doctors and this system from March 1, 2021 (week 9) to January 14, 2022 (week 2) were collected, as well as weekly positive rate of influenza virus (WPRIV) released by China National Influenza Center. Eight weekly absenteeism rate indicators were calculated: all-cause absenteeism rate reported by system (WAR1), all-cause absenteeism rate reported by school doctors (WAR2), sickness absenteeism rate reported by school doctors (WAR3), and the rate of students absent one time (WAR4), two times (WAR5), three to four times (WAR6), one to two times (WAR7) and two to four times (WAR8) a week reported by the system. Pearson correlation coefficients of eight indicators and WPRIV were analyzed, and the change trend of their time series diagram was investigated. *Results* During week 9-42, WAR1 ($r=0.614$, $p=0.001$), WAR4 ($r=0.631$, $p<0.001$), WAR5 ($r=0.651$, $p<0.001$), WAR6 ($r=0.541$, $p<0.001$), WAR7 ($r=0.654$, $p<0.001$) and WAR8 ($r=0.644$, $p<0.001$) were significantly correlated with WPRIV, while WAR2 ($r=0.262$, $p>0.05$) and WAR3 ($r=0.239$, $p>0.05$) were not. Throughout the surveillance period, WAR1 ($r=0.671$, $p<0.001$), WAR2 ($r=0.638$, $p<0.001$), WAR3 ($r=0.752$, $p<0.001$), WAR5 ($r=0.682$, $p<0.001$), WAR6 ($r=0.535$, $p<0.001$) and WAR8 ($r=0.683$, $p<0.001$) were significantly correlated with WPRIV, while WAR4 ($r=0.086$, $p>0.05$) and WAR7 ($r=0.242$, $p>0.05$) were not. *Conclusions* Absenteeism reported by the system was more effective for influenza surveillance than absenteeism reported by school doctors, especially when the influenza activity level was low. When WAR1, WAR5 and WAR8 were combined together, the epidemic situation of influenza could be more comprehensively aware.

Keywords: Face Recognition, Absenteeism, Influenza, Syndromic Surveillance

1. Introduction

Schools play an important role in the spread of influenza [1]. School-age children have weak immune systems, and the schools where they spend most of their time are crowded. This makes school-age children vulnerable to influenza and to spreading it to others. According to statistics, a child with influenza can directly infect more than 2.4 classmates around him [2]. Children with flu can again carry the infection to their families, allowing it to spread into the wider community [3]. Adults who live with school-age children have a two to three times higher risk of influenza than those who do not live with children [4].

Influenza surveillance among school-age children is an important measure to control influenza. Absenteeism has become an important indicator of school-based influenza symptom surveillance since a 1979 study by Peterson and colleagues verified that student absenteeism is highly correlated with influenza epidemic levels in the community [5]. The advantages of absenteeism surveillance include non-invasive, no need for clinical tests, low cost, simple operation and good representation, helping to accurately assess the economic burden of infectious diseases and their impact on education, and promoting effective collaboration between the health and education sectors [6-11]. Absenteeism is classified as all-cause absenteeism, sickness absenteeism and symptom-specific absenteeism [12]. All-cause absence involves confounding factors such as personal leave or accidental injury, but sickness absenteeism and symptom-specific absenteeism can remove partial bias. Consequently, Donaldson et al. [12] conclude that symptom-specific absenteeism has more advantages in influenza syndromic surveillance. Although Crawford et al. [9] pointed out that all-cause absenteeism could properly reflect the epidemic situation of influenza in the community only when the epidemic level was high. Baer et al. [10] believed that of its simplicity, all-cause absenteeism was also an effective indicator of influenza surveillance when resources were inadequate.

Sickness absenteeism and symptom-specific absenteeism are nearly used as surveillance indicators in the existing school-based infectious disease syndromic surveillance system in China [13-18]. This mode of operation relies heavily on school doctors to collect information on various symptoms, thus enhancing the positive predictive value of surveillance while also imposing a heavier work burden on schools, particularly for larger schools. Nevertheless, only 33.1% of primary and secondary schools in China are equipped with school doctors, and on average, a school doctor needs to serve over 2800 people [19]. The huge workload makes it difficult for school doctors to timely and accurately report absenteeism information, and many absenteeism surveillance systems have reported data quality problems. The average school utilization rate of the system in Hangzhou was only 54.13% [15], 40.56% of cases in the system in Shanghai were not reported in time [16], and the accuracy and timeliness of the system data report in Xi 'an was only 72.32% and 83.93% [14]. Absenteeism surveillance must strike a balance between specificity and school burden, otherwise it cannot give full play to the due value of surveillance [10]. The development of absenteeism surveillance in China urgently needs mode innovation.

There have been attempts to replace manual absenteeism statistics with fingerprint scans [20] or smart cards [21]. These two methods require frequent human-machine contact, which will bring a high risk of cross infection. In this study, an absenteeism statistical method based on face recognition was proposed, which achieves contactless and intelligent absenteeism collection and analysis. A primary school with 1900 students was selected as a pilot school. The effect of different methods on influenza surveillance was evaluated by comparison between the absenteeism reported by the new system and by the school doctor, then a surveillance index system suitable for the new system was constructed. The results of this study would provide a new solution to the development dilemma of absenteeism surveillance.

2. Methods

2.1. Reporting System

Xixaolianxing is a campus management APP based on the Alipay platform. After the school signs an agreement with the operating company, the school organizes the parents to download the APP and register an account in Alipay for free through their smart phones. According to the agreement, parents require to input the child's name, gender, ID, school name, class, face image and other information into the account. This information will be uploaded to the APP's data processing center and stored in the student identity database at different levels (individual, class, and school). The data acquisition terminal of APP is several face recognition devices (SUNMI-FT1Mini, recognition accuracy ≥ 99.99%). Devices are generally installed at the school gate. When students arrive at the school gate in the morning, they must go through the instrument identification before entering the school. If a student is not tested within one hour of the school's scheduled attendance deadline, he will be counted as absent. Daily absenteeism and attendance information will be classified by different levels, and then feed back to different users such as parents and schools.

2.2. Study Population

This study selected a primary school in Hangzhou, Zhejiang Province, which was equipped with a school doctor. School doctor began to collect absenteeism information as required by their superiors in March 2020. The school began to utilize the new APP in November 2020, and the system operation data of the school from March 1, 2021 to January 14, 2022 were extracted. The surveillance period comprised two phases: Phase I, From March 1,2021 to June 25,2021, the effective surveillance period was 83 days. During this period, the school had 357, 393, 345, 240, 248, 258 students from grade 1 to grade 6, with a total of 1,861 students; Phase II, From September 1,2021 to January 14, 2022, the effective surveillance period was 91 days. During this period, the school had 317, 383, 404, 357, 249, 250 students from grade 1 to Grade 6, with a total of 1960 students. In both phases, all of the school's students signed up for accounts on the APP.

2.3. Date Collection

This study collected three types of data: (1) Information of absentee students reported by the system, which was exported from the background of the system, including daily absentee name, class, school and other information; (2) Information of absent students reported by the school doctor, which was collected by the school doctor every day, including the name, class, school and reason (or symptoms) of absent students all day; (3) Weekly Positive Rate of Influenza Virus (WPRIV), which was downloaded from the official website of Chinese National Influenza Center (CNIC, https://ivdc.chinacdc.cn/cnic/). WPRIV is equal to the ratio of weekly virus-positive samples to the total number of samples submitted for testing in each region, which is divided into the south and the north. Hangzhou is located in the south of China, so this study only refers to data from the south. Corresponding to the start time of this study, WPRIV was collected from the

6th week of 2021 to the 5th week of 2022. During this period, influenza virus strains prevalent in southern China were mainly type B. Because we used only de-identified data, the Tongji University Review Board designated this study as nonhuman subjects research.

2.4. Data Analysis

First, we calculated the daily all-cause absenteeism rate (DAR1) reported by the system based on the system export information. Then the daily all - cause absence rate (DAR2) and daily sickness absence rate (DAR3) reported by school doctors were calculated based on the data. In order to investigate the correlations of the three variables, we drew their time series diagrams and calculated their Pearson correlation coefficients. The calculation formula of the three indicators was as follows:

DAR1 = (Daily number of all-cause absence reported by the system/ Number of students enrolled in school) * 100%

DAR2 = (Daily number of all-cause absence reported by the doctor / Number of students enrolled in school) * 100%

DAR1 = (Daily number of sickness absence reported by the doctor / Number of students enrolled in school) * 100%

Secondly, based on daily absenteeism, we calculated the following eight variables of weekly absence rate (WAR): all-cause absenteeism rate reported by the system (WAR1), all-cause absenteeism rate reported by the school doctor (WAR2), sickness absenteeism rate reported by the school doctor (WAR3), and the rate of students absent one time (WAR4), two times (WAR5), three to four times (WAR6), one to two times (WAR7) and two to four times (WAR8) a week reported by the system. The calculation formula of the eight variables was as follows:

WAR1 = [Weekly number of all-cause absence reported by the system / (M*N)]* 100%

WAR2 = [Weekly number of all-cause absence reported by the doctor / (M*N)]* 100%

WAR3 = [Weekly number of sickness absence reported by the doctor / (M*N)]* 100%

WAR4 = [Number of students absent one time a week reported by the system / (M*N)]* 100%

WAR5 = [Number of students absent two times a week reported by the system / (M*N)]* 100%

WAR6 = [Number of students absent 3-4 times a week reported by the system / (M*N)]* 100%

WAR7 = [Number of students absent 1-2 times a week reported by the system / (M*N)]* 100%

WAR8 = [Number of students absent 2-4 times a week reported by the system / (M*N)]* 100%

In these formulas, M represents the number of students enrolled in school, and N represents effective surveillance days of the week.

Finally, we draw a time series diagram of the eight week absenteeism rate indicator and WPRIV to intuitively compare the effects of different indicators on influenza surveillance. At the same time, the correlation and timeliness of different indicators in predicting influenza epidemic were investigated by analyzing the Pearson correlation coefficient between eight week absenteeism rate and WPRIV under four conditions: no advance (t), advance one week (t-1), advance two weeks (t-2) and advance three weeks (t-3).

3. Results

3.1. Analysis of DAR

The results showed (Table 1) that the values of the three DAR indicators were lower in phase I and higher in phase II. According to the definition, the absenteeism reported by school doctors was a subset of the absenteeism reported by the system, and the sickness absence was included in the all-cause absence reported by school doctors. In phase I, the system reported 2475 absenteeism and the school doctor reported 414 absenteeism (of which 341 were due to illness). The number of absenteeism reported by the doctor accounted for 16.73% of the number of absenteeism reported by the system. The number of sickness absence accounted for 82.37% of the school doctors reported absence and 13.78% of the system reported absence. In phase II, the system reported 4770 absenteeism, school doctors reported 1476 absenteeism, of which 322 were due to COVID-19 isolation and 976 were due to illness. The number of absenteeism reported by school doctors accounted for 30.94% of the number of absenteeism reported by the system. The number of sickness absence accounted for 66.12% of the number of all-cause absence reported by school doctor and 20.46% of the number of all-cause absence reported by the system.

In phase I, DAR1 & DAR2 (r=0.201, p=0.077) and DAR1 & DAR3 (r=0.125, p=0.266) were not significantly correlated, while DAR2 and DAR3 (r=0.937, p=0.000) were significantly positively correlated. In phase II, DAR1 & DAR2 (r=0.779, p=0.000), DAR1 & DAR3 (r=0.548, p=0.000), and DAR2 & DAR3 (r=0.830, p=0.000) were all highly positively correlated. After summarizing the data of the two phases, the three pairs of variables of DAR1& DAR2 (r=0.740, p=0.000), DAR1 & DAR3 (r=0.578, p=0.000) and DAR2 & DAR3 (r=0.855, p=0.000) were also highly positively correlated.

Table 1. *The descriptive statistics of the daily all-cause absenteeism rate reported by the system, the daily all-cause absenteeism rate and the daily sickness absenteeism rate reported by school doctor.*

Variables	Phase I				Phase II			
	Min	Max	Mean	SD	Min	Max	Mean	SD
The daily all-cause absenteeism rate reported by system (%)	0.86	4.19	1.64	0.582	1.07	5.63	2.54	0.955
The daily all-cause absenteeism rate reported by doctor (%)	0.00	1.13	0.28	0.200	0.00	4.40	0.82	0.885
The daily sickness absenteeism rate reported by doctor (%)	0.00	1.07	0.22	0.176	0.00	1.84	0.54	0.413

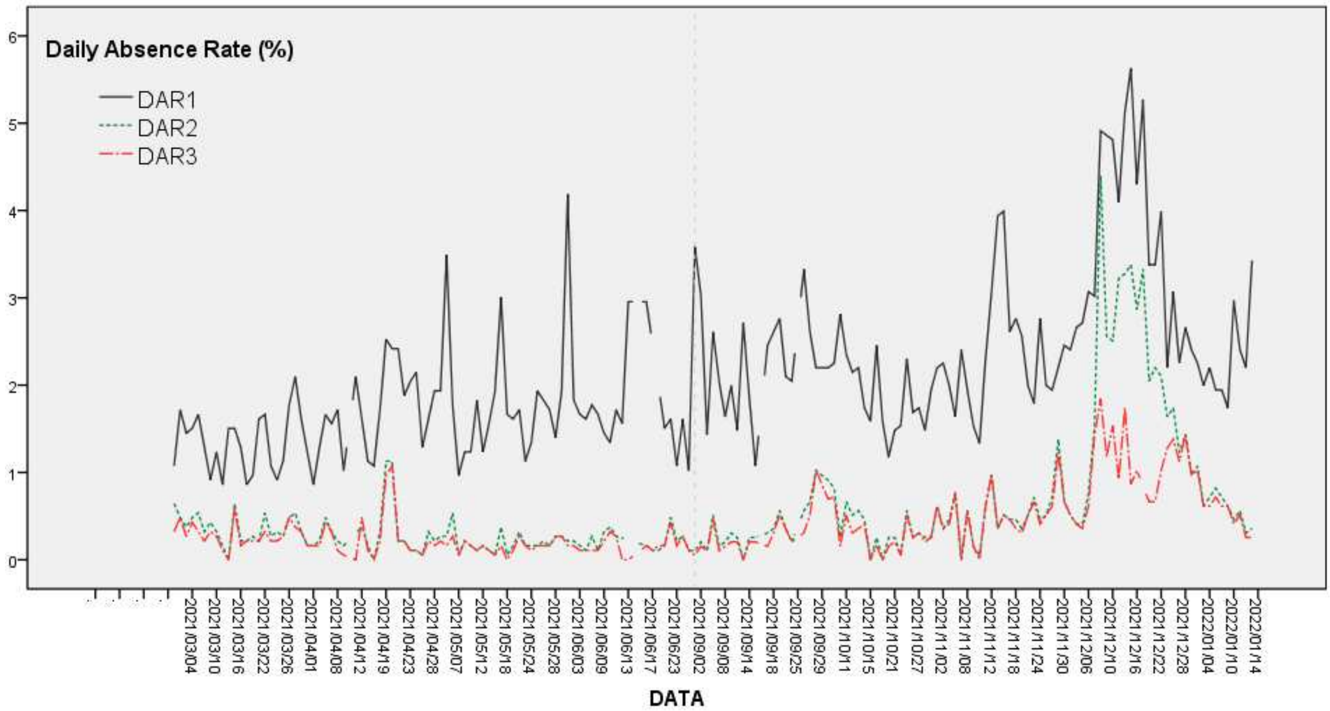

Figure 1. *Time series of DAR1, DAR2 and DAR3.*

Note: DAR1, the daily all-cause absenteeism rate reported by the system; DAR2, the daily all-cause absenteeism rate reported by school doctor; DAR3, the daily sickness absenteeism rate reported by school doctor.

As can be seen from the time series (Figure 1), DAR1, DAR2 and DAR3 had comparatively high consistency in curve changes, particularly in the second phase. On April 19, September 26, and December 10, 2021, DAR1, DAR2, and DAR3 curves will peak simultaneously. Nevertheless, at the end of the second stage, DAR1 curve had a warping phenomenon, while DAR2 and DAR3 did not. The distance between DAR2 and DAR3 curves was very near, and the variation trend was highly consistent. However, there were significant differences between the two curves during the three weeks from December 6 to December 24, 2021, during which some students were isolated due to the COVID-19 epidemic.

Table 2. *The correlation matrix of WAR1, WAR2, WAR3, WAR4, WAR5, WAR6, WAR7 and WAR8.*

Variables	WAR1	WAR2	WAR3	WAR4	WAR5	WAR6	WAR7	WAR8
WAR1	1.000							
WAR2	0.811***	1.000						
WAR3	0.714***	0.899***	1.000					
WAR4	0.406*	0.032	0.011	1.000				
WAR5	0.697***	0.500***	0.551***	0.532**	1.000			
WAR6	0.690***	0.625***	0.650***	0.541***	0.545***	1.000		
WAR7	0.516**	0.150	0.145	0.978***	0.657***	0.611***	1.000	
WAR8	0.783***	0.677***	0.714***	0.563***	0.768***	0.956***	0.691***	1.000

Note: (1)* p<0.05, ** p<0.01, *** p<0.001. (2) WAR1, weekly all-cause absenteeism rate reported by system; WAR2, weekly all-cause absenteeism rate reported by doctor; WAR3, weekly sickness absenteeism rate reported by doctor; WAR4, weekly all-cause absenteeism (one time a week) rate reported by system; WAR5, weekly all-cause absenteeism (two times a week) rate reported by system; WAR6, weekly all-cause absenteeism (3-4 times a week) rate reported by system; WAR7, weekly all-cause absenteeism (1-2 times a week) rate reported by system; WAR8, weekly all-cause absenteeism (2-4 times a week) rate reported by system.

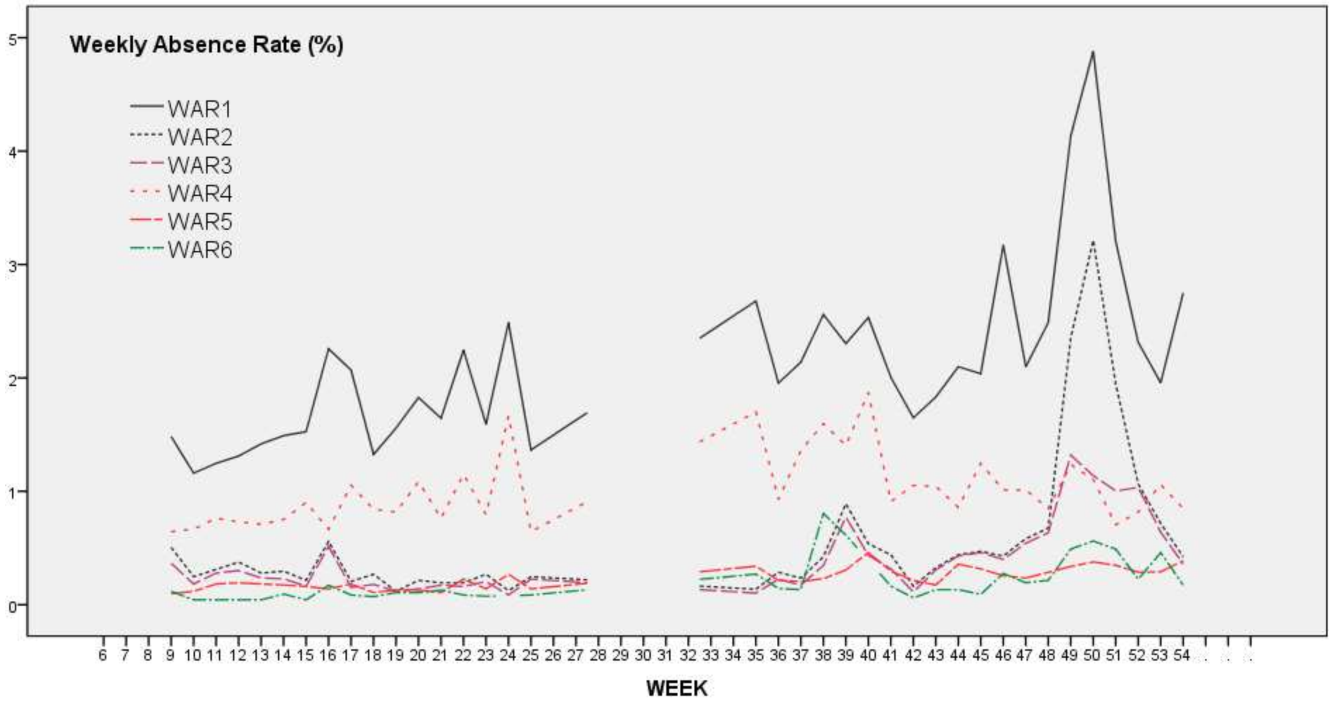

Figure 2. *Time series of WAR1, WAR2, WAR3, WAR4, WAR5 and WAR6.*

Note: WAR1, weekly all-cause absenteeism rate reported by system; WAR2, weekly all-cause absenteeism rate reported by doctor; WAR3, weekly sickness absenteeism rate reported by doctor; WAR4, weekly all-cause absenteeism (one time a week) rate reported by system; WAR5, weekly all-cause absenteeism (two times a week) rate reported by system; WAR6, weekly all-cause absenteeism (3-4 times a week) rate reported by system.

3.2. Analysis of WAR

In phase I, there were 1225 students (77.88%) absent once a week, 227 students (14.43%) absent twice a week, and 121 students (7.69%) absent 3-4 times a week. In phase II, there were 1963 students (65.67%) absent one a week, 517 students (17.54%) absent twice a week, and 495 students (16.79%) absent 3-4 times a week. Hence, the majority of absenteeism is once a week. In contrast, the frequency of three types of absenteeism rose significantly in the second phase, and the absenteeism frequency of twice per week and three to four times per week increased more significantly. The time series diagram (Figure 2) revealed that WAR1 and WAR4 had a high degree of trend coincidence before week 42, and the other four curves were very adjacent. After the 42nd week, only WAR4 was in a downward channel, while the other five curves expanded in different amplitude after a temporary convergence appeared in the 42nd week.

In order to further explore the correlation between the data reported by the system and the data reported by school doctors, Pearson correlation coefficients of eight weekly absenteeism rate indicators were calculated (Table 2). If using WAR3 as a reference, WAR2 had the highest correlation, followed by WAR1 and WAR8, and WAR6 was the third. The correlation between WAR4, WAR7, and WAR3 was not significant, and even these three indicators have only a moderately positive correlation with WAR1. However, if we only count the data before 42 weeks, the correlation coefficients of WAR1 and WAR4, WAR7 and WAR8 were 0.853 (p=0.000), 0.852 (p=0.000), and 0.745 (p=0.000), respectively.

3.3. Correlation Between WAR and WPRIV

We counted the correlation coefficients between eight weekly absenteeism rate indicators and WPRIV (Table 3). When the level of influenza was low (week 9-42), WAR2 and WAR3 were not significantly correlated with WPRIV, while WAR1, WAR4, WAR5, WAR6, WAR7, and WAR8 were significantly correlated with WPRIV, and WAR5 had the highest correlation with WPRIV. After adding data from 43 to 54 weeks (with high flu levels), the correlation between WAR2 and WAR3 and WPRIV became significant, and the correlation between WAR3 and WPRIV was the largest among all the coefficients. WAR1, WAR5, WAR6, and WAR8 were still significantly correlated with WPRIV, while WAR4 and WAR7 were not significantly correlated with WPRIV. Among them, the correlation coefficient between WAR1 and WAR5 and WPRIV was the most prominent.

The time series diagram of WAR1, WAR3, WAR5, WAR8 and WPRIV demonstrated that before 42 weeks (WPRIV was at a low level), the curves of WAR1, WAR5 and WAR8 fit well with the curve of WPRIV, while the curve of WAR3 fit badly with the curve of WPRIV. From week 43 to 54 (WPRIV was at a high level), the time series curves of WAR1, WAR5, WAR6, and WAR8 had a high degree of coincidence with the time series curve of WPRIV. Throughout the process, WAR1 stands out. On the 23rd week, WPRIV experienced the peak, while WAR1 displayed the stage peak at the 22nd and 24th weeks; the time series curve of WPRIV showed peaks at week

43, 46, 50, and 53, and the four peaks increased successively, among which only WAR1 had the highest similarity.

Table 3. *Correlation analysis of eight week absenteeism rate index and WPRIV under four conditions.*

Time	Variables	WPRIV			
		t	t-1	t-2	t-3
9-42 weeks	WAR1	0.614***	0.597***	0.560***	0.521***
	WAR2	0.262	0.172	0.198	0.216
	WAR3	0.239	0.151	0.164	0.179
	WAR4	0.631***	0.606***	0.558***	0.581***
	WAR5	0.651***	0.638***	0.639***	0.682***
	WAR6	0.541***	0.475*	0.458*	0.431*
	WAR7	0.654***	0.631***	0.591***	0.618***
	WAR8	0.644***	0.595**	0.586**	0.549**
9-54 weeks	WAR1	0.671***	0.646***	0.666***	0.694***
	WAR2	0.638***	0.620***	0.642**	0.670***
	WAR3	0.752***	0.747***	0.760**	0.778***
	WAR4	0.086	0.081	0.070	0.115
	WAR5	0.682***	0.683***	0.672***	0.677***
	WAR6	0.535***	0.510***	0.522***	0.512***
	WAR7	0.242	0.238	0.225	0.265
	WAR8	0.683***	0.666***	0.673***	0.658***

Note: (1)* p<0.05, ** p<0.01, *** p<0.001. (2) WAR1, weekly all-cause absenteeism rate reported by system; WAR2, weekly all-cause absenteeism rate reported by doctor; WAR3, weekly sickness absenteeism rate reported by doctor; WAR4, weekly all-cause absenteeism (one time a week) rate reported by system; WAR5, weekly all-cause absenteeism (two times a week) rate reported by system; WAR6, weekly all-cause absenteeism (3-4 times a week) rate reported by system; WAR7, weekly all-cause absenteeism (1-2 times a week) rate reported by system; WAR8, weekly all-cause absenteeism (2-4 times a week) rate reported by system; WPRIV, weekly positive rate of influenza virus.

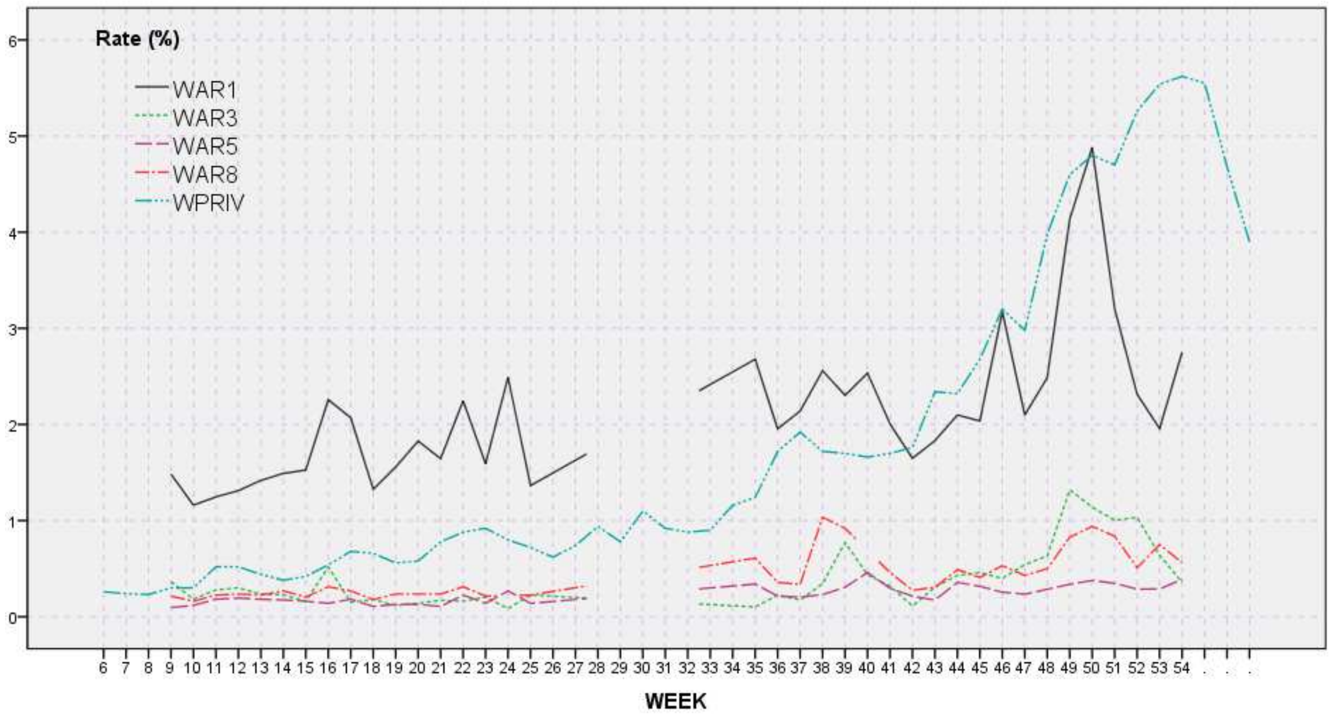

Figure 3. *Time series of WAR1, WAR3, WAR5, WAR8 and WPRIV.*

Note: (1) We made WPRIV 5 times smaller for better graphics. (2) WAR1, weekly all-cause absenteeism rate reported by system; WAR2, weekly all-cause absenteeism rate reported by doctor; WAR3, weekly sickness absenteeism rate reported by doctor; WAR4, weekly all-cause absenteeism (one time a week) rate reported by system; WAR5, weekly all-cause absenteeism (two times a week) rate reported by system; WAR6, weekly all-cause absenteeism (3-4 times a week) rate reported by system; WAR7, weekly all-cause absenteeism (1-2 times a week) rate reported by system; WAR8, weekly all-cause absenteeism (2-4 times a week) rate reported by system; WPRIV, weekly positive rate of influenza virus.

4. Discussion

There are four main findings in this study: (1) Absenteeism automatically collected by face recognition was accurate and reliable; (2) The effect of influenza surveillance was significantly affected by the absence duration, and the short-term absence was better than long-term absence; (3) In view of China's national conditions, all-cause absenteeism may be a better indicator at present; (4) The comprehensive application of multiple indicators based on the new system can better perceive the epidemic situation of influenza.

Data quality is the key to the effectiveness of surveillance systems [22]. The system collects absenteeism data through face recognition, which not only reduces the school's burden, but also greatly improves data timeliness and accuracy. There are three reasons for this: (1) During the surveillance period, all the students in the sentinel school registered their accounts on the APP, and the system's coverage of the objects was theoretically complete; (2) The face recognition technology of the system comes from Alipay, its security and accuracy are of financial level, and the probability of students arriving at school without being recognized is extremely low; (3) The school has arranged a guard to supervise the operation of devices, and only students whose identities are confirmed by devices are permitted to enter the school, so it is highly unlikely that they will arrive at school without being tested by devices. The absenteeism reported by school doctors has official endorsement, so its quality is guaranteed. As a reference standard, the absenteeism reported by the system was highly positively correlated with the standard, and the change trend of their time series curves was also highly consistent. Accordingly, it is feasible to collect absenteeism through our surveillance system, but the necessary operational supervision is indispensable.

The delineation of absence time had a significant effect on influenza surveillance. Absences were defined by the system as "not being tested by device within one hour of the school arrival deadline", and by the school doctor as "not being in school that day". Definitely, the former is a short-term absence, while the latter is a long-term absence, and the latter is a subset of the former. When influenza activity levels were low, the absenteeism reported by school doctors was only 16.73% of the absenteeism reported by the system, and when influenza activity levels increased, this figure increased to 30.94%. Which one is a reflection of the truth? In China, the high emphasis on education makes it difficult for parents to let their children leave the classroom easily. According to a survey [13], only 37.41% of the students who saw a doctor because of illness would ask for leave, that is, 62.59% of the students would still go to school after seeing a doctor because of illness. Those who return to school after seeing a doctor are not counted in the data reported by school doctors, but they can be partially monitored by the new system. The results showed that at low levels of influenza activity, short-term absence was significantly associated with influenza activity while long-term absence was not. When influenza activity levels increased, both long and short-term absence were significantly correlated with influenza activity levels, but the latter had a greater correlation. The prevalence of school attendance with illness is considered to be one of the main reasons for the poor absenteeism surveillance effect in China [23], and short-term absenteeism may have a better elimination effect on this bias than long-term absenteeism.

Sickness absence is generally considered to be more specific than all-cause absence [11, 12]. Sickness absence can exclude some false positive cases of absence such as personal leave or accidental injury, so its surveillance specificity is theoretically superior to that of all-cause absence. The data in this study also supported this view. The correlation coefficient between WAR3 and WPRIV was 0.752, while that between WAR2 and WPRIV was 0.638. In phase I, the number of sickness absence accounted for 82.37% of the all-cause absence reported by the school doctor, and the correlation coefficient between them was 0.937. In phase II, the corresponding figures were 66.12% (84.58% after excluding isolated students) and 0.855, respectively. School doctors reported sickness absence and all - cause absence were highly consistent, and there was not much difference in the correlation between them in WPRIV. Especially when influenza activity level was low, only WAR2 and WAR3 were not significantly correlated with WPRIV. Statistics of sickness absence require more school medical resources, but the current prevalence rate of school doctors in China is about 33.1% [19]. In addition, sickness absence was not closely related to WPRIV when the influenza level was low, so the advantage of sickness absence may not be prominent in China, and all-cause absence could be a better indicator for influenza surveillance.

Schmidt et al. suggested that the prevalence of absenteeism is better than the incidence of absenteeism in terms of influenza surveillance effectiveness [8]. The findings of this study helped to extend our understanding of this conclusion. WAR4 was similar to incidence and the other seven-week absenteeism indicators were similar to prevalence. When the level of influenza activity was moderate, WAR4 was significantly positively correlated with WPRIV, while it was negatively correlated with WPRIV when the level of influenza activity was high. Consequently, overall, WAR4 was not significantly correlated with WPRIV. WAR1, WAR5 and WAR8 were positively correlated with WPRIV regardless of high or low level of influenza activity. These three indicators may reflect the different status of influenza illness among students: WAR5 represents more cases of moderate influenza, WAR8 represents severe influenza, and WAR1 represents the complete spectrum from mild to severe influenza. Therefore, the comprehensive application of WAR1, WAR5 and WAR8 indicators can help us better understand the epidemic situation of influenza.

5. Limitations

This study explored the feasibility of collecting and analyzing students' absenteeism data through face recognition, compared the difference of the effects of different absenteeism indicators in influenza surveillance, and provided a new solution for the optimization of absenteeism monitoring. However, there are some limitations in this study, mainly in two aspects: (1) Insufficient research samples, limited by conditions this study only investigated the data of one school with a sample size of about 1,900 people, but this is not enough, and the applicability of our research conclusions in other primary and secondary schools has not been fully verified; (2) COVID-19 caused some data deviations. No COVID-19 cases occurred in sentinel schools during the implementation of this study, but the absenteeism data collected were distorted due to the pressure of epidemic prevention and control. For example, the isolation of some students led to the exaggerated absenteeism rate. In the future, we need to use a larger sample size in non-epidemic periods to verify whether the conclusions of this study are still applicable.

6. Conclusions

The absenteeism statistic method based on face recognition is feasible, which is contactless and data collection is automatic. According to the analysis, the absenteeism data measured by the standard of system (not tested within one hour of the school's scheduled attendance deadline), is more valuable for influenza surveillance than that measured by the standard of school doctor (absent all day). The following three indicators were closely correlated with the level of influenza prevalence: all-cause absenteeism rate reported by system (WAR1), the rate of students absent two times (WAR5) and two to four times (WAR8) a week reported by the system. The combination of these three indicators can make a more comprehensive awareness of the influenza epidemic situation.

Acknowledgements

This study is indebted to the National Influenza Center of China for providing influenza epidemic data, as well as the sentinel schools and partner companies for their strong support of this study. Most importantly, the Ji 'an Science and Technology Bureau provided financial support for this study.

References

[1] W. R. Dowdle, J. D. Millar, L. B. Schonberger, et al. "Influenza immunization policies and practices in Japan," J Infect Dis. 1980, vol. 141 (2), pp 258-264.

[2] Y. Yang, J. D. Sugimoto, M. E. Halloran, et al. "The transmissibility and control of pandemic influenza A (H1N1) virus," Science, 2009, vol. 326 (5953), pp 729-733.

[3] I. M. Longini, J. S. Koopman, A. S. Monto, and J. P. Fox. "Estimating household and community transmission parameters for influenza," Am J Epidemiol, 1982, vol. 115 (5), pp 736-751.

[4] J. Woodall, K. E. Rowson, J. C. McDonald. "Age and Asian influenza, 1957," BMJ, 1958, vol. 2 (5108), pp 1316-1318.

[5] D. Peterson, J. S. Andrews, B. S. Levy, et al. "An effective school-based influenza surveillance system," Public Health Rep, 1979, Vol. 94 (1), pp 88-92.

[6] D. D. Lenaway, A. Ambler. "Evaluation of school-based influenza surveillance systems," Public health reports, 1995, Vol. 110 (3), pp 333-337.

[7] P. Mook, C. Joseph, P. Gates, and N. Phin. "Pilot scheme for monitoring sickness absence in schools during the 2006/07 winter in England: can these data be used as a proxy for influenza activity?" Euro Surveill, 2007, Vol. 12 (12), pp E11-E12.

[8] W. P. Schmidt, R. Pebody, P. Mangtani. "School absence data for influenza surveillance: a pilot study in the United Kingdom," Euro Surveill, 2010, Vol. 15 (3), ppii=19467.

[9] G. B. Crawford, S. Mckelvey, J. Crooks, et al. "Influenza and school- based influenza-like illness surveillance: a pilot initiative in Maryland," Public Health Report, 2011, Vol. 126, pp 591-596.

[10] A. Baer, C. V. Rodriguez, J. S. Duchin. "Anautomated system for public health surveillance of school absenteeism," J Public Health Management Practice, 2011, Vol. 17 (1), pp 59-64.

[11] V. L. Short, C. K. Marriott, S. Ostrof, and K. Walle. "Description and evaluation of the 2009-2010 Pennsylvania Influenza Sentinel School Monitoring System," American Journal of Public Health, 2011, Vol. 101 (11), pp 2178-2183.

[12] A. L. Donaldson, J. L. Hardstaff, J. P. Harris, et al. "School-based surveillance of acute infectious disease in children: a systematic review," BMC Infectious Diseases, 2021, Vol. 21, pp 744.

[13] C. J. Pan, Q. Zhao, C. M. Zhou, et al. "The influence of sickness absence of elementary school students on the effectiveness of absence monitoring of infectious disease surveillance system in rural area in Jiangxi," Chin Prev Me, 2014, Vol. 15 (8), pp 705-709.

[14] H. Zhang, X. T. Fan, S. J. Zhang, et al. "Analysis on quality of school straight report system of symptoms monitoring absence because of illness in Xi'an City," J Med Pest Control, 2019, Vol. 35 (10), pp 942-945.

[15] H. D ing, C. P. Huang, B. Wang, et al. "A smartphone-based school syndromic surveillance system and its application in Hangzhou," Chin J Sch Health, 2019, Vol. 40 (12), pp 1862-1864.

[16] Y. H. Chen, Y. H. Wang, W. Wang, et al. "Early warning effect of school absence monitoring on prevention and control of infectious diseases," Chin J Sch Health, 2020, Vol. 41 (3), pp 465-467.

[17] X. Dong, M. L. Boulton, B. Carlson, et al. "Syndromic surveillance for influenza in Tianjin, China: 2013–14," Journal of Public Health, vol. 39 (2), pp 274-281.

[18] L. Tan, L. Cheng, W. Yan, et al. "Using daily syndrome-specific absence data for early detection of school outbreaks: a pilot study in rural China," public health, vol. 128 (2014), pp 792-798.

[19] H. Z. Zhou, G. R. Zhu, X. Zhang, et al. "Current situation and analysis of school physicians in primary and secondary schools in 16 provinces in China," Chin J Sch Health, 2018, Vol. 39 (10), pp 1455-1458.

[20] Saranath Lawpoolsri, Amnat Khamsiriwatchara, Wongwat Liulark, et al. "Real-Time Monitoring of School Absenteeism to Enhance Disease Surveillance: A Pilot Study of a Mobile Electronic Reporting System," JMIR mHealth uHealth, 2014, Vol. 2 (2), pp e22.

[21] K. M. Dennis, E. H. Y. Lau, C. S. Hau, et al. “A Smart Card-Based Electronic School Absenteeism System for Influenza-Like Illness Surveillance in Hong Kong: Design, Implementation, and Feasibility Assessment.,” JMIR Public Health Surveill, 2017, Vol. 3 (4), pp e67.

[22] S. L. Groseclose, D. L. Buckeridge. “Public health surveillance systems: recent advances in their use and evaluation,” nnu. Rev. Public Health, 2017, Vol. 38, pp 57-79.

[23] X. Gao, H. B. Ma. “Analysis on the absence of primary and secondary school students due to illness in Xicheng District, Beijing,” Chin J Sch Health, 2007, Vol. 28 (11), pp 1037-1038.